

(19)



JAPANESE PATENT OFFICE

PATENT ABSTRACTS OF JAPAN

(11) Publication number: **11085193 A**(43) Date of publication of application: **30.03.99**

(51) Int. Cl.

G10L 5/04**G10L 3/00****G10L 9/18**(21) Application number: **09248750**(22) Date of filing: **12.09.97**(71) Applicant: **SANYO ELECTRIC CO LTD**(72) Inventor:
HIRAI HIROYUKI
ONISHI HIROKI
NISHIDA HIDEJI
HASHIMOTO MAKOTO(54) **PHONEME INFORMATION OPTIMIZATION
METHOD IN SPEECH DATA BASE AND
PHONEME INFORMATION OPTIMIZATION
APPARATUS THEREFOR**

processing section 5 executes the clustering processing to the number assigned by using the LBG algorithm in the case the distribution is known in accordance with the calculated selection probability.

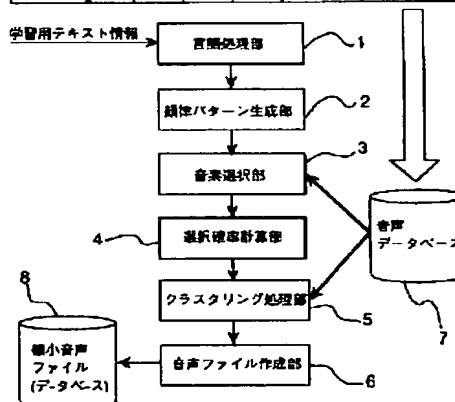
(57) Abstract:

COPYRIGHT: (C)1999,JPO

PROBLEM TO BE SOLVED: To form a speech file from phoneme information of high use frequencies by determining the use frequencies of each pieces of phoneme information constituting a speech data base and executing clustering processing in accordance with these use frequencies.

SOLUTION: A rhythm pattern forming section 2 estimates the basic frequency F_0 near the phoneme center, power and phoneme duration time by using the part-of- speech information of the input text obtainable from the results of phoneme symbol, accent symbol string and morphological analyses. Next, a cost is determined and the phonemes are selected. The text information (sentence) for learning is synthesized and the number of the selected times of the respective phonemes is calculated by using the speech data base 7 including all the phonemes. Next, the number of the selected times of the phonemes included in the respective phoneme units is averaged to form the number of selected times of the phoneme units and the selection probability of the respective phoneme units is calculated. A clustering

波形データ	ラベル	ピッチ	パワー	時間量	ケプストラム	カウンタ
134.05~203.75	/a n/	130	6321	63.8	1.8, -1.2, ..., -0.12	29
203.75~274.2	/n u/	228	6347	68.4	2.1, -0.2, ..., -0.02	65
...



特開平11-85193

(43) 公開日 平成11年(1999) 3月30日

(51) Int. Cl. ⁶ 識別記号G10L 5/04
3/00
9/18

F I

G10L 5/04 E
3/00 H
9/18 E

審査請求 未請求 請求項の数 9 O L (全 8 頁)

(21) 出願番号 特願平9-248750

(22) 出願日 平成 9 年(1997) 9 月12日

(71) 出願人 000001889

三洋電機株式会社

大阪府守口市京阪本通 2 丁目 5 番 5 号

(72) 発明者 平井 啓之

大阪府守口市京阪本通 2 丁目 5 番 5 号 三
洋電機株式会社内

(72) 発明者 大西 宏樹

大阪府守口市京阪本通 2 丁目 5 番 5 号 三
洋電機株式会社内

(72) 発明者 西田 秀治

大阪府守口市京阪本通 2 丁目 5 番 5 号 三
洋電機株式会社内

(74) 代理人 弁理士 安富 耕二 (外 1 名)

最終頁に続く

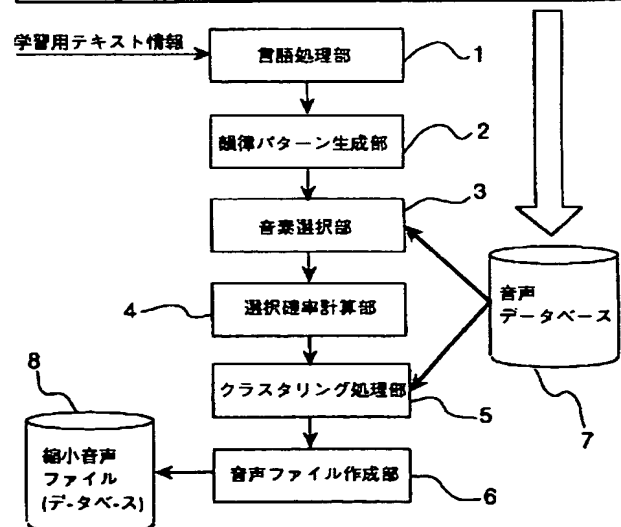
(54) 【発明の名称】 音声データベースにおける音素片情報最適化方法、及び音素片情報最適化装置

(57) 【要約】

【課題】 従来の音声データベース最適化方法によってクラスタリング処理して音素片情報を削減したとしても、コンテキストクラスタテーブルには音声合成に際して全く使用されない音素片情報を多く含んだままの状態であるといった問題があった。

【解決手段】 本発明は、文章発話から切り出した音素片を接続することにより合成音を得る波形合成に適用される音声データベースにおける音素片情報最適化方法において、予め学習用テキスト情報を用いて合成し、その合成結果に従って前記音声データベースを構成する各音素片情報の使用頻度を求め、該使用頻度に基づいてクラスタリング処理を行うことにより、音声ファイルの音素片情報を最適化する。

波形データ	ラベル	ピッチ	パワー	時間長	ケプストラム	カウンタ
134.95~203.75	/e n/	190	6621	68.8	1.6, -1.2, ..., -0.12	30
203.75~271.2	/n o/	226	6347	68.4	2.1, -0.2, ..., -0.02	65
:	:	:	:	:	:	:



【特許請求の範囲】

【請求項 1】 文章発話から切り出した音素片を接続することにより合成音を得る波形合成に適用される音声データベースにおける音素片情報最適化方法において、予め学習用テキスト情報を用いて合成し、その合成結果に従って前記音声データベースを構成する各音素片情報の使用頻度を求め、該使用頻度に基づいてクラスタリング処理を行うことにより、音声ファイルの音素片情報を最適化することを特徴とする音声データベースにおける音素片情報最適化方法。

【請求項 2】 文章発話から切り出した音素片を接続することにより合成音を得る波形合成に適用される音声データベースにおける音素片情報最適化方法において、音素選択部が、学習用テキスト情報を入力として、前記文章発話から切り出した音素片を蓄積した音声データベースから最適な音素片を選択する第 1 ステップと、選択確率計算部が、前記音素選択部によって選択された各音素片の選択確率を求める第 2 ステップと、クラスタリング処理部が、前記音声データベースに対し、所定のパラメータ空間において、前記選択確率を音素片の分布確率としてクラスタリング処理を行う第 3 ステップと、及び音素波形素片登録部が、前記クラスタリング処理部によってクラスタリングされた、各クラスタの中から代表音素片を選択する第 4 ステップ、からなることを特徴とする音声データベースにおける音素片情報最適化方法。

【請求項 3】 文章発話から切り出した音素片を接続することにより合成音を得る波形合成に適用される音声データベースにおける音素片情報最適化装置において、前記文章発話から切り出した音素片を蓄積した音声データベースと、学習用テキスト情報を入力として、前記音声データベースからなる最適な音素片を選択する音素選択部と、該音素選択部によって選択された、各音素片の選択確率を求める選択確率計算部と、前記音声データベースに対し、所定のパラメータ空間において、前記選択確率を音素片の分布確率としてクラスタリング処理を行うクラスタリング処理部と、該クラスタリング処理部によってクラスタリング処理された、各クラスタの中から代表音素片を選択する音素波形素片登録部と、を備えることを特徴とする音声データベースにおける音素片情報最適化装置。

【請求項 4】 前記クラスタリング処理部は、各クラスタ内のセントロイドから前記クラスタ内に含まれる全ての音素片までの距離が最小になるようにクラスタリング処理することを特徴とする請求項 1、又は 2 記載の音声データベースにおける音素片情報最適化方法。

【請求項 5】 前記学習用テキスト情報は、文章から構成されていることを特徴とする請求項 1、又は 2 記載の音声データベースにおける音素片情報最適化方法。

【請求項 6】 前記音素片は、少なくとも波形情報から構成されていることを特徴とする請求項 1、又は 2 記載の音声データベースにおける音素片情報最適化方法。

【請求項 7】 前記クラスタリング処理部は、各クラスタ内のセントロイドから前記クラスタ内に含まれる全ての音素片までの距離が最小になるようにクラスタリング処理することを特徴とする請求項 3 記載の音声データベースにおける音素片情報最適化装置。

【請求項 8】 前記学習用テキスト情報は、文章から構成されていることを特徴とする請求項 3 記載の音声データベースにおける音素片情報最適化装置。

【請求項 9】 前記音素片は、少なくとも波形情報から構成されていることを特徴とする請求項 3 記載の音声データベースにおける音素片情報最適化装置。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は、予め文章発話から切り出して蓄積した、音素片情報からなる音声データベースから最適な音素片情報を選択し接続することにより合成音を得る波形合成に適用される、音声データベースにおける音素片情報最適化方法、及び音素片情報最適化装置に関する。

【0002】

【従来の技術】従来、音声波形を接続して合成音を得る波形合成に適用される音声データベースの音素片情報に対してクラスタリングを行い、最適な音素片情報を音声ファイルに登録する音声ファイル構成方式等が特開平 8 - 2 6 3 5 2 0 号公報に開示されている。

【0003】図 5 は、従来のコンテキストクラスタリングの処理を示すフローチャートである。同図において、音声データベース 100 内の音素ラベリングされた波形データ中から同一の音素ラベルが付与されている波形データを全て取り出し、初期クラスタ 110 とする（ステップ 201）。

【0004】次に、この初期クラスタ 110 内の個々の波形データ（要素）を特徴分析する（ステップ 202）。この特徴分析においては、LPC（線形予測符号化法）ケプストラム等の特徴パラメータの次数を n とし、かつ、分析窓関数のフレーム周期を可変として、フレーム数が m フレームとなるように分析を行うことにより、各要素に対して $n \times m$ 次元の特徴パラメータ行列を得る。

【0005】次にこの特徴分析の結果を用いて、初期クラスタ 110 のクラスタ歪を求める（ステップ 203）。具体的には、特徴パラメータのベクトル空間において、初期クラスタ 110 の全ての要素と予め求めておいたセントロイドとの間の距離の 2 乗和を求めて、これを初期クラスタ 110 のクラスタ歪と定義する。

【0006】こうして初期クラスタ 110 のクラスタ歪を求め、これをコンテキストクラスタテーブル 208 に

登録する。このコンテキストクラスタテーブル 208 には、図示のように、各クラスタ毎に、それに属するコンテキストと、そのセントロイドと、そのクラスタ歪と、それに含まれる要素波形の集合とが登録されている。

【0007】尚、初期クラスタ 110 のクラスタ歪を求めた段階では、初期クラスタ 100 だけがコンテキストクラスタテーブル 208 に登録されていることになる。

【0008】次にコンテキストクラスタテーブル 208 中からクラスタ歪が最大となるクラスタを求め（ステップ 204）、この求めたクラスタを、コンテキストクラスタテーブル 208 中から取り出し、コンテキストにより更に 2 つのクラスタに分割する（ステップ 205）。

【0009】尚、最初の段階では、初期クラスタ 110 だけがコンテキストクラスタテーブル 208 に登録されているので、この初期クラスタ 110 に対してクラスタ分割が行われる。

【0010】このようにして、初期クラスタ 110 の分割が行われた後、コンテキストクラスタテーブル 208 において、初期クラスタ 110 が削除され、分割された 2 つのクラスタが新たに登録される（ステップ 206）。

【0011】以上の処理（ステップ 203～206）を繰り返すことにより、初期クラスタ 110 は次第に小さいクラスタに細分化されていく。そして、この各繰り返しループ毎に、コンテキストクラスタリングの終了判定が行われる（ステップ 207）。

【0012】

【発明が解決しようとする課題】然し乍ら、この音声データベース 100 の音素片情報を削減して音声ファイル（データベース）を作成したとしても、音声データベース 100 に含まれる文章と音声合成器に入力する文章とでは音素片の出現頻度が異なるため、コンテキストクラスタテーブル 208 には音声合成に際して全く使用されない音素片情報を多く含んだままの状態であるといった問題が依然残っていた。

【0013】従って、本発明は、大量の学習用テキスト情報（文章）を予め用意し、それを全ての音素片を用いた音声合成器で予め合成し、その結果から各音素片の使用された回数（頻度情報）を求め、その分布にしたがって距離の総和を計算し、クラスタリングを行うことを特徴とする。

【0014】これによって、クラスタリング処理の対象となっている音声データベースに様々な音素片情報が含まれていたとしても、頻繁に使用される音声に対して多くの音素片を割り当てた音声ファイル（データベース）を構築することが可能となる。

【0015】

【課題を解決するための手段】本発明の音声データベースにおける音素片情報最適化方法は、文章発話から切り出した音素片を接続することにより合成音を得る波形合

成に適用される音声データベースにおける音素片情報最適化方法において、予め学習用テキスト情報を用いて合成し、その合成結果に従って前記音声データベースを構成する各音素片情報の使用頻度を求め、該使用頻度に基づいてクラスタリング処理を行うことにより、音声ファイルの音素片情報を最適化することを特徴とする。

【0016】また、本発明の音声データベースにおける音素片情報最適化方法は、文章発話から切り出した音素片を接続することにより合成音を得る波形合成に適用される音声データベースにおける音素片情報最適化方法において、音素選択部が、学習用テキスト情報を入力として、前記文章発話から切り出した音素片を蓄積した音声データベースから最適な音素片を選択する第 1 ステップと、選択確率計算部が、前記音素選択部によって選択された各音素片の選択確率を求める第 2 ステップと、クラスタリング処理部が、前記音声データベースに対し、所定のパラメータ空間において、前記選択確率を音素片の分布確率としてクラスタリング処理を行う第 3 ステップと、及び音素波形素片登録部が、前記クラスタリング処理部によってクラスタリングされた、各クラスタの中から代表音素片を選択する第 4 ステップ、からなることを特徴とする。

【0017】本発明の音声データベースにおける音素片情報最適化装置は、文章発話から切り出した音素片を接続することにより合成音を得る波形合成に適用される音声データベースにおける音素片情報最適化装置において、前記文章発話から切り出した音素片を蓄積した音声データベースと、学習用テキスト情報を入力として、前記音声データベースからなる最適な音素片を選択する音素選択部と、該音素選択部によって選択された、各音素片の選択確率を求める選択確率計算部と、前記音声データベースに対し、所定のパラメータ空間において、前記選択確率を音素片の分布確率としてクラスタリング処理を行うクラスタリング処理部と、該クラスタリング処理部によってクラスタリング処理された、各クラスタの中から代表音素片を選択する音素波形素片登録部と、を備えることを特徴とする。

【0018】また、前記クラスタリング処理部は、各クラスタ内のセントロイドから前記クラスタ内に含まれる全ての音素片までの距離が最小になるようにクラスタリング処理することを特徴とする。

【0019】前記学習用テキスト情報は、文章から構成されていることを特徴とする。

【0020】更に、前記音素片は、少なくとも波形情報から構成されていることを特徴とする。

【0021】

【発明の実施の形態】本発明の実施の形態を図 1～図 4 を用いて説明する。

【0022】図 1 は、本発明を実現するための装置の概略構成図である。また、図 2 は、本発明における、音声

データベースにおける音素片情報最適化方法を実現するためのフローチャートである。

【0023】以下、図1を参照し乍ら、図2の処理過程を説明する。

【0024】ステップS1では、学習用テキスト情報（文章）が言語処理部1に入力されると、言語処理部1は、形態素解析、係り受け解析を行い、解析後の音素に対して音素記号、品詞、及びアクセント記号列を付与する。

【0025】ステップS3では、韻律パターン生成部2は音素記号、アクセント記号列、及び形態素解析の結果より得られる入力テキストの品詞情報を用いて、音素中心付近での基本周波数 F_0 、パワー、音韻継続時間長を

$$D(F) = \sum_{i=1}^k (w_{F0} D_{F0}(c(i)) + w_{pow} D_{pow}(c(i)) + w_{dur} D_{dur}(c(i)) + w_{posi} D_{posi}(c(i))) + \max(w_{F0}^c D_{F0}^c(c(i)) + w_{pow}^c D_{pow}^c(c(i)) + w_{cep}^c D_{cep}^c(c(i)) + w_{ph}^c D_{ph}^c(c(i)))$$

【0028】尚、 D_{F0} 、 D_{pow} 、 D_{dur} は、音素中心付近での基本周波数、パワー、音韻継続時間長の推定値と合成単位との差であり、 D_{posi} は、文中の位置（語頭、語中、及び語尾）の違いを数値化した値である。

D_{F0}^c 、 D_{pow}^c 、 D_{cep}^c は、接続する2つの合成単位の接続点での基本周波数の差、パワーの差、ケプストラムの差である。 D_{ph}^c は、発話環境を考慮して決定された接続の行い易さ（接続優先順位）を示す歪である。また、 w_i 、 w_i^c は、夫々のパラメータに乗ずる重み係数である。

【0029】次に、ステップS7では、全ての音素片を含む音声データベース7を用いて、学習用テキスト情報（文章）を合成し、各音素片の選択された回数を計算する。具体的には、音声データベース7の全ての音素片を、適当な音素単位に分割する。このとき、無声の子音を含む場合には、CV、VCに分割し、有声の子音を含む場合には、VCVに分割している。尚、「C」とは、子音（Consonant）を表わし、また「V」とは、母音（Vowel）を表わす。

$$D(F) = \sum_{i=1}^k (w_{F0} D_{F0}(c(i)) + w_{pow} D_{pow}(c(i)) + w_{dur} D_{dur}(c(i)) + w_{posi} D_{posi}(c(i)))$$

【0035】ステップS25では、各クラスに属する全ての音素片に関する重心（セントロイド）を計算し、 $m+1$ の代表ベクトルとする。このセントロイドは、音素片の音響パラメータのベクトルの各要素ごとの平均を

推定する。

【0026】ステップS5においては、数1に示す式を用いてコストを求め、音素を選択する。本ステップにおける具体的な音素片の選択は、ステップS3で推定された基本周波数 F_0 、パワー、音韻継続時間長の推定値との非適合を表わすコスト、及び各音素片を接続するときのコストを計算し、その総和が最小になる音素片の組み合わせをDP（ダイナミックプログラミング）法に従い数1を用いることにより探索を行う。ここで、コストを示すコスト関数 $D(F)$ を数1に示す。

【0027】

【数1】

【0030】次に、各音素単位に含まれる音素片の選択された回数を平均し、音素単位の選択回数とし、各音素単位の選択確率を計算する。尚、本発明では、1度も選択されなかった音素単位にも小さな確率を割り当てることとした。ステップS9においては、ステップS7で計算された選択確率に基づいて、クラスタリング処理部5は、分布が既知の場合のLBGアルゴリズムを用いて、指定された個数にクラスタリング処理を行う。

【0031】ここで、ステップS9を図3を用いて詳細に説明する。

【0032】まず、ステップS21では、指定された個数の初期代表ベクトル A_0 を任意に決定する。またインデックス $m=0$ 、平均歪み $D_{-1}=\infty$ とする。

【0033】ステップS23では、音声データベースの全ての音素片を最も近い代表ベクトル A_m が属するクラス $P(A_m)$ に分割する。この時の距離の計算は数2を用いる。

【0034】

【数2】

計算することで求められるが、この平均は、各音素の選択確率を用いて計算される。

【0036】ステップS27では、代表ベクトル A_{m+1} 、クラス $P(A_{m+1})$ の時の平均歪み D_{m+1} を計算

する。歪みは、前記数 2 を選択確率で平均した結果である。

【0037】ステップ S 29 は、インデックスを 1 増加させる。

【0038】ステップ S 30 は、終了判定を行っている。歪みの減少率を計算し一定量 ϵ 以下ならその時のクラス $P(A_s)$ を出力として終了する。

【0039】ここで、図 2 に戻って更に説明を続ける。

【0040】最終的に、ステップ S 11 では、ステップ S 9 で求められた各クラス P のセントロイドを計算し、それに最も近い音素を選択音素として音声ファイル作成部 6 が登録することによって、縮小 (削減) された音声ファイル (データベース) が新たに作成される。

【0041】次に、本発明の有効性を確かめるため、評価実験を行った。本実験では、地名の読み上げを行う合成器の生成を目的とした。学習用文章には、新郵便番号

データのうち九州地方を除く全てを用いた。新郵便番号データより、市・郡名称、区町村名称、町域名称を抽出し、「ここは、X 市、Y 区、Z 町、です。」という文章に変換し合成を行った。

【0042】その結果より、「ここは、」と「です。」の部分を除き、残りの結果より各音素の選択確率を求めた。求めた選択確率を用いてクラスタリングした縮小ファイル (データベース) と、選択確率が同様としてクラスタリングした縮小ファイル (データベース) を用いて、学習に用いた地名、学習に用いなかった地名 (九州地方)、小説の 3 種類の文章を合成し評価した。

【0043】以下に実験に用いた音声ファイル (データベース) のサイズ、および実験結果を示す。

【0044】

【表 1】

学習用テキスト	305745 単語、2635765 音素
音声データベース	48069 音素
縮小データベース	5500 音素

【0045】図 4 において、縦軸は、地名 20 文章、小説 5 文章を合成した時の数 1 の歪コストの合計を文章の総音素数で割った 1 音素当りの平均歪である。また、図 4 中の斜線は選択確率が同様として作成した音声ファイル (データベース) による合成結果 (conventional)、また交差線は提案方式による結果 (proposed)、更に縦線は全ての音素片を含む音声ファイル (データベース) による合成結果 (all) である。

【0046】place-name (closed) は学習に用いた地名、place-name (open) は学習に用いなかった地名 (九州地方)、novel は全く環境の異なる文章である小説の結果を示す。

【0047】この結果より、全ての場合で提案方式の方が選択確率を同様とした場合と比較して歪が少なくなっており、提案方式が有効であることがわかる。それぞれの文章の種類ごとに比較すると、proposed の歪は地名読み上げでは all に近いが、小説読み上げでは conventional に近い。これは、open-closed に関わらず言えることで、地名読み上げという環境への最適化が行われていることがわかる。

【0048】

【発明の効果】以上の説明から明らかなように、本発明によれば、文章発話から切り出した音素片を接続することにより合成音を得る波形合成に適用される音声データベースにおける音素片情報最適化方法において、予め学習用テキスト情報を用いて合成し、その合成結果に従って前記音声データベースを構成する各音素片情報の使用頻度を求め、該使用頻度に基づいてクラスタリング処理

を行うことにより、使用頻度の高い音素片情報からなる音声ファイル (データベース) を作成することができる効果を奏する。

【0049】更に、本発明は、文章発話から切り出した音素片を接続することにより合成音を得る波形合成に適用される音声データベースにおける音素片情報最適化装置において、前記文章発話から切り出した音素片を蓄積した音声データベースと、学習用テキストを入力として、前記音声データベースからなる最適な音素片を選択する音素選択部と、該音素選択部によって選択された、各音素片の選択確率を求める選択確率計算部と、前記音声データベースに対し、所定のパラメータ空間において、前記選択確率を音素片の分布確率としてクラスタリング処理を行うクラスタリング部と、該クラスタリング部によってクラスタリングされた、各クラス P の代表音素片を選択する音素波形素片登録部と、を備えることにより、使用頻度の高い音声には多くの音素片情報を割り当てることが出来る効果を奏する。

【図面の簡単な説明】

【図 1】本発明を実現するための装置の概略構成図である。

【図 2】本発明における、音声データベースにおける音素片情報最適化方法を実現するためのフローチャートである。

【図 3】図 2 に示すステップ S 9 の処理を詳細に表したフローチャートである。

【図 4】本発明の評価実験の結果を示す図である。

【図 5】従来のコンテキストクラスタリングの処理を示す

すフローチャートである。

【符号の説明】

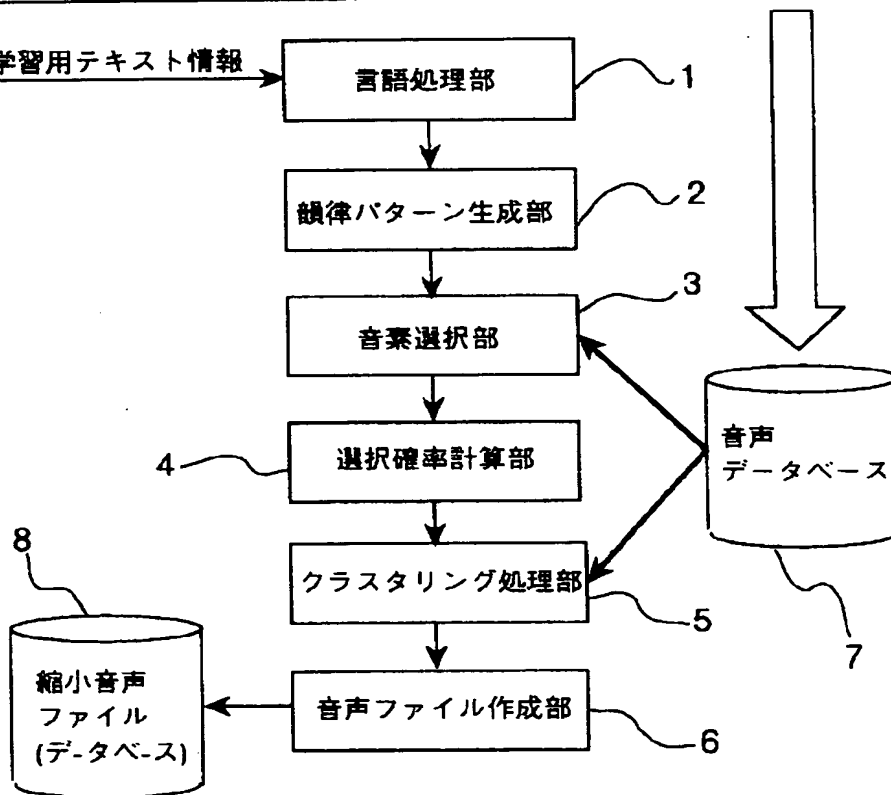
- 1 …… 言語処理部
2 …… 韻律パターン生成部
3 …… 音素選択部

- 4 …… 選択確率計算部
5 …… クラスタリング処理部
6 …… 音声ファイル作成部
7 …… 音声データベース
8 …… 縮小音声ファイル (データベース)

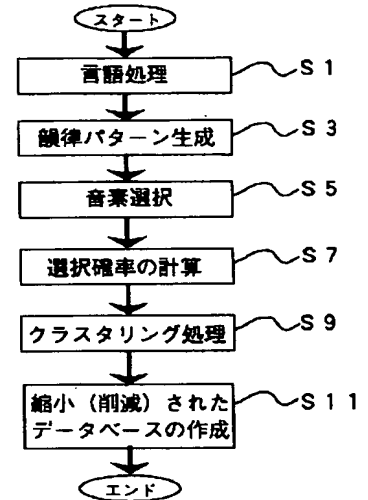
【図 1】

波形データ	ラベル	ピッチ	パワー	時間長	ケプストラム	カウンタ
134.95~203.75	/ a n /	190	6621	68.8	1.6, -1.2, …… , -0.12	30
203.75~271.2	/ n o /	226	6347	68.4	2.1, -0.2 …… , -0.02	65
⋮	⋮	⋮	⋮	⋮	⋮	⋮

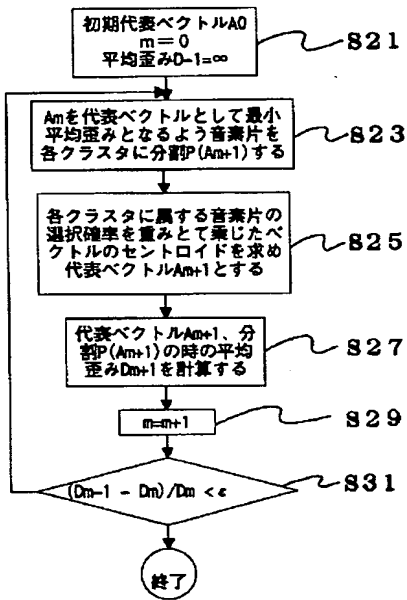
学習用テキスト情報



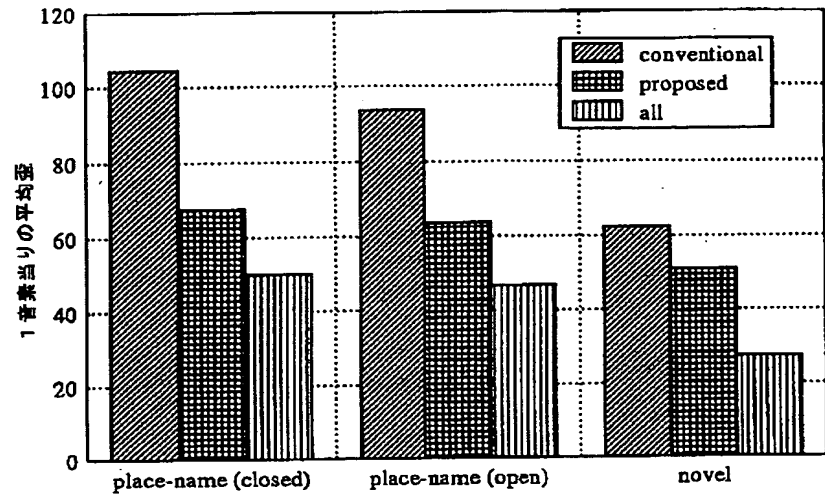
【図 2】



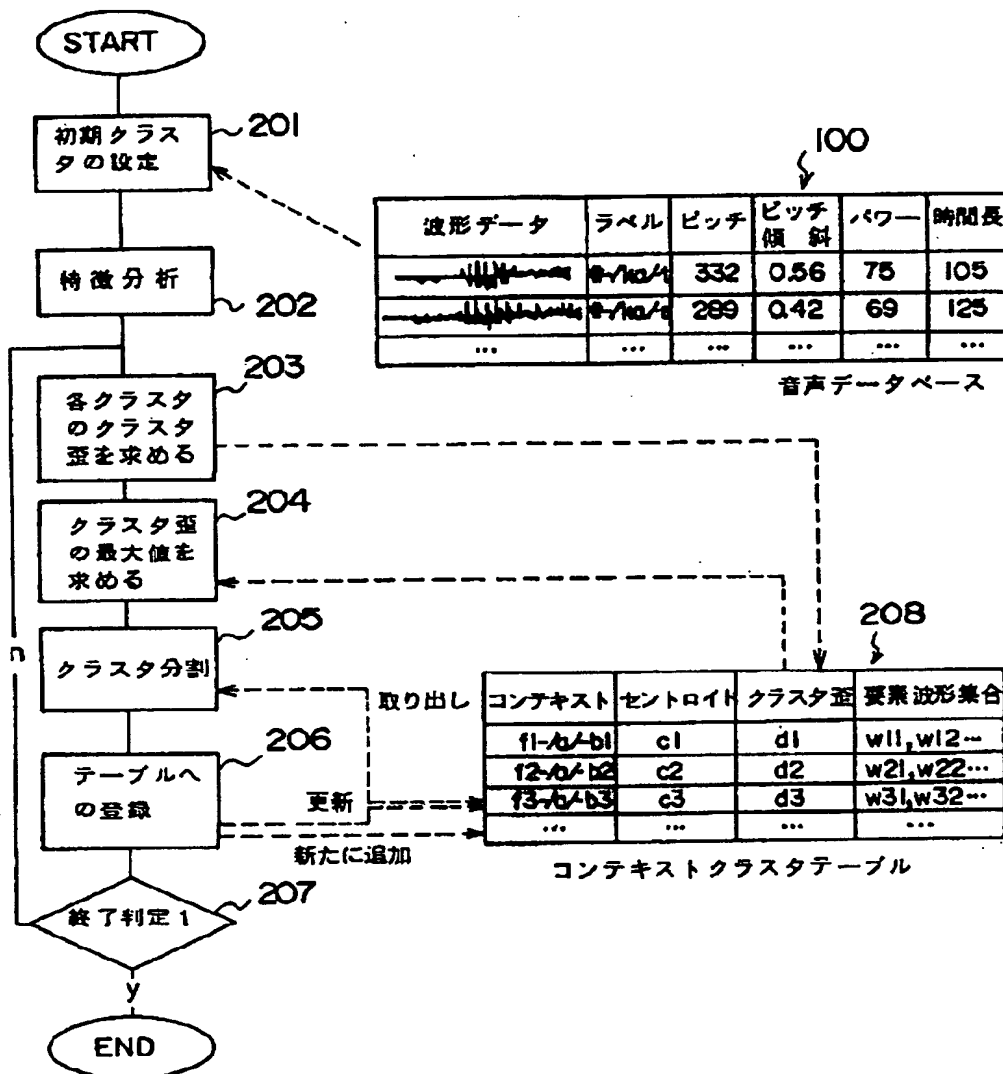
【図 3】



【図 4】



【図 5】



フロントページの続き

(72)発明者 橋本 誠

大阪府守口市京阪本通 2 丁目 5 番 5 号 三
洋電機株式会社内